

Dynamic Graph Message Passing Networks

Li Zhang

Department of Engineering Science
University of Oxford
lz@robots.ox.ac.uk

Dan Xu

Department of Engineering Science
University of Oxford
danxu@robots.ox.ac.uk

Anurag Arnab

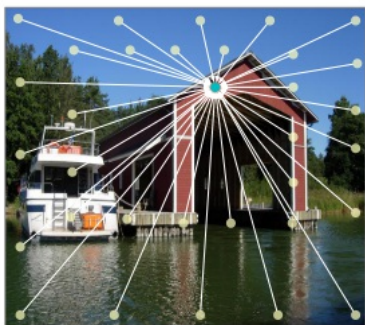
Department of Engineering Science
University of Oxford
aarnab@robots.ox.ac.uk

Philip H.S. Torr

Department of Engineering Science
University of Oxford
phst@robots.ox.ac.uk

目的：

根据输入动态采样特征图信息，建立高效率的信息传输路径与动态权重，实现高效的信息传输。通常的卷积运算是局部连接的图，而 non-local 是全连接的图，二者可能都会包含冗余的信息传输。本文方法降低空间注意力的计算复杂度，也实现了动态的感受野。



(a) Fully-connected message passing



(b) Locally-connected message passing



(c) Dynamic graph message passing

方法：

信息的传输过程实际上就是特征图中的点进行加权组合。位置 i 的信息的计算：

算：

$$\mathbf{m}_i^{(t+1)} = M^t \left(A_{i,j}, \{\mathbf{h}_1^{(t)}, \dots, \mathbf{h}_K^{(t)}\}, \mathbf{w}_j \right) = \sum_{j \in \mathcal{N}(i)} A_{i,j} \mathbf{h}_j^{(t)} \mathbf{w}_j, \text{ with } A_{i,j} = A[i, j]$$

矩阵 A 表示点与点连接强度； \mathbf{w} 是特征变换矩阵，可由卷积实现； \mathbf{h} 为特征

图上的某个点。

文中方法的**动态**包含两个方面：对不同位置的点使用不同的点进行加权；以及加权的权重根据输入动态调整。

具体地：对于特征图中的某个点，用不同参数的平均分布在特征图上采样 S 组，每组包含 K 个点。然后对这 S 组点分别做一次随机游走，根据特征来计算平移的量：

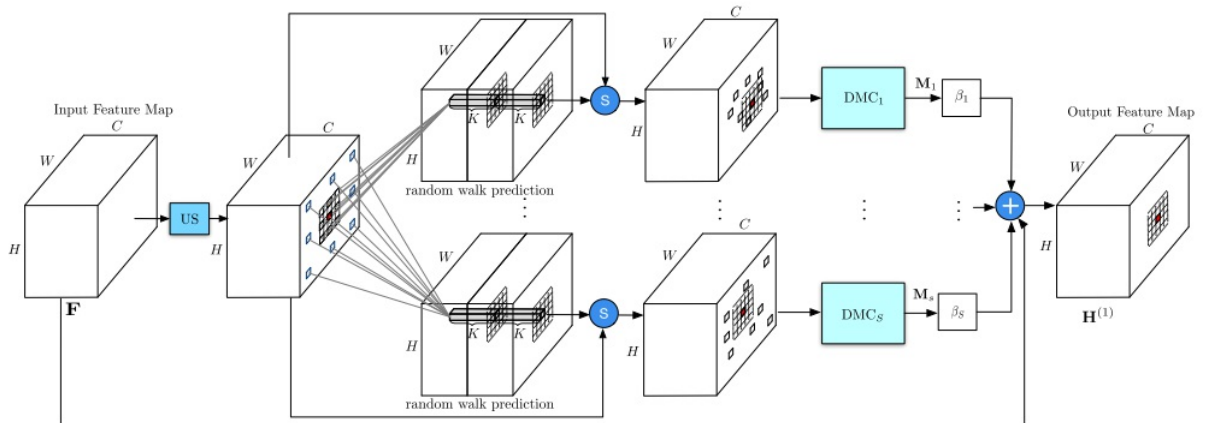
$$\Delta \mathbf{d}_j^q = \mathbf{W}_{i,j}^q v_i^q + \mathbf{b}_{i,j}^q,$$

\mathbf{v} 表示某一组采样得到的 K 个点的特征。然后对均匀采样的点集 \mathbf{v} 中的每个点在特征图上做不同的平移。

因为点可能游走到非网格点，所以还要对非网格点做个双线性差值，然后用新的点去计算信息。然后对这 S 组点计算得到的信息再加权平均，得到最终的信息。

$$\mathbf{m}_i^{(t+1)} = \sum_q \sum_{j \in \mathcal{N}_q(i)} \beta_q A'^q_{i,j} \varrho \left(\mathbf{h}_j^{(t)} | \mathcal{V}, j, \Delta \mathbf{d}_j^q \right) \mathbf{w}_j^q, \text{ with } A'^q_{i,j} = A'^q[i, j],$$

到这里,我们知道了对于特征图中的某个点,要从周围 $S \times K$ 个点来收集信息。



接下来是如何动态计算信息，即上图中的青色方框部分。

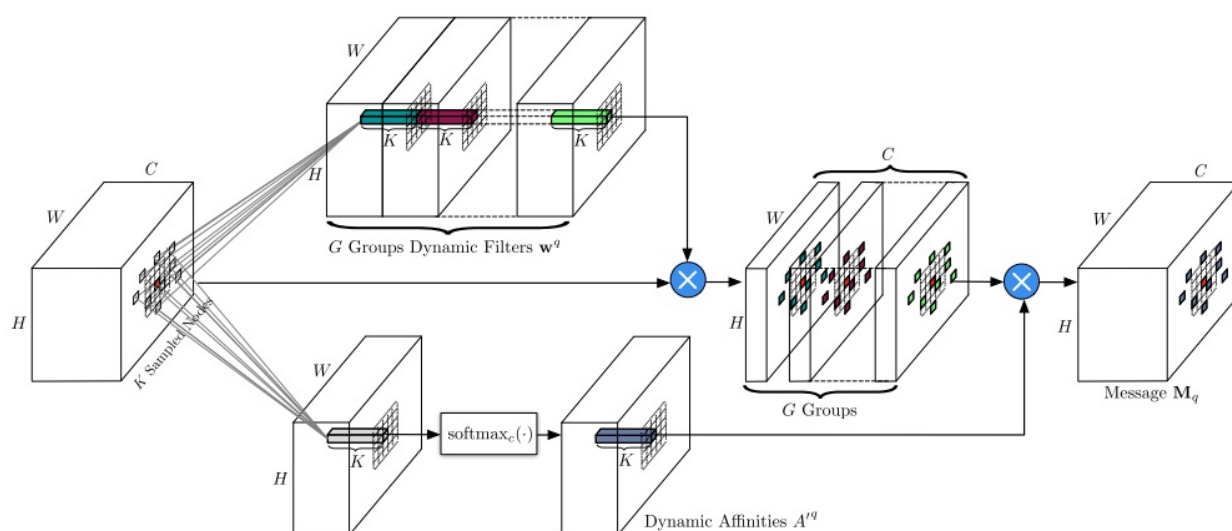
每组采样得到的数据又使用不同的连接强度（用矩阵 A 表示），且 K 个参考像素各自对应不同的变换 w

$$\{\mathbf{w}_j^q, A'_{i,j}^q\} = \mathbf{W}_{i,j}^{k,A} v_i'^q + \mathbf{b}_{i,j}^{k,A},$$

$$A'_{i,j}^q = \text{softmax}_c(A'_{i,j}^q) = \frac{\exp(A'_{i,j}^q)}{\sum_{l \in \mathcal{N}_q(i)} \exp(A'_{i,l}^q)},$$

使用点集中 K 个点的特征来计算变换矩阵和连接强度矩阵，对连接强度另做 softmax，使得 K 个点的权重之和=1。

为了节约参数，还将将通道 C 分为 G 个组，同一个组中的通道使用相同的参数。



结论：

DGMN 的计算复杂度降低了，且精度超过了 non-local。我觉得主要原因是 non-local 建立的是所有点之间的联系，对特征图中的点都是使用相同的卷积运算；而 DGMN 每个点都动态地使用不同的变换，由于只从特征图中的若干点收集信息，所以也不会产生太多的参数和计算量的增加，相比之下更加地灵活。